

Asymptotics of Transmit Antenna Selection: Impact of Multiple Receive Antennas

Saba Asaad^{*†}, Ali Bereyhi[†], Ralf R. Müller[†], Amir M. Rabiei^{*}

^{*}School of Electrical and Computer Engineering, University of Tehran

[†]Institute for Digital Communications (IDC), Friedrich-Alexander Universität Erlangen-Nürnberg (FAU)

saba_asaad@ut.ac.ir, ali.bereyhi@fau.de, ralf.r.mueller@fau.de, rabiei@ut.ac.ir

Abstract— Consider a fading Gaussian MIMO channel with N_t transmit and N_r receive antennas. The transmitter selects L_t antennas corresponding to the strongest channels. For this setup, we study the distribution of the input-output mutual information when N_t grows large. We show that, for any N_r and L_t , the distribution of the input-output mutual information is accurately approximated by a Gaussian distribution whose mean grows large and whose variance converges to zero. Our analysis depicts that, in the large limit, the gap between the expectation of the mutual information and its corresponding upper bound, derived by applying Jensen's inequality, converges to a constant which only depends on N_r and L_t . The result extends the scope of channel hardening to the general case of antenna selection with multiple receive and selected transmit antennas. Although the analyses are given for the large-system limit, our numerical investigations indicate the robustness of the approximated distribution even when the number of antennas is not large.

I. INTRODUCTION

Massive Multiple-Input Multiple-Output (MIMO) systems have recently received a great deal of interest due to their promise of high performance gains [1]. These gains are mainly achieved at the expense of having a tremendous number of antenna elements within a relatively small physical platform. From practical points of view, moving to millimeter wave spectrum can make this issue conceivable [2]. The growth in the number of antennas, however, increases the Radio Frequency (RF) cost significantly. Therefore, addressing solutions to alleviate this issue has become a topic of interest. Antenna selection is a possible solution which reduces hardware costs dramatically without significant performance loss [3], [4]. For MIMO channels with limited number of antennas, the performance under different measures, such as capacity, Signal-to-Noise Ratio (SNR) at the receiver, bit error rate, and outage probability has been investigated in the literature [3]. There are, however, few results which have addressed the large-system analysis of antenna selection [5], [6], and the asymptotic behavior for a general MIMO setup is still unknown.

Asymptotic Channel Hardening

The asymptotic hardening property of MIMO fading channels was first studied in [7]. The property indicates that in the large-system limit, the distribution of the mutual information between a white Gaussian input and the output of a MIMO Gaussian fading channel concentrates almost normally around its mean while the variance shrinks rapidly. Under single

Transmit Antenna Selection (TAS), the channel hardening was studied initially in [4], where the authors considered a TAS protocol selecting a single transmit antenna with the strongest channel gain. It was further shown that, under this scheme, the channel hardens at a slower rate compared to the case considered in [7]. In [5], the limiting behavior of the mutual information in an uplink channel was investigated considering the transmitter to be equipped with a single transmit antenna, and the receiver to select a number of strongest channels. For this scenario, the distribution of the input-output mutual information was approximated with the distribution of the logarithm of a folded normal random variable. In the large-system limit, it was further shown that the variance converges to zero which concludes the asymptotic hardening property for the setup. The asymptotics of TAS were further studied in [6] for a downlink scenario with single antenna receiver in which a multi-antenna transmitter selects a number of antennas with strongest channel coefficients. For this scenario, the input-output mutual information was approximated asymptotically, and the hardening property was shown to hold.

In this paper, we generalize the earlier studies by determining a large-system approximation for the input-output mutual information of a Gaussian MIMO channel, when both the transmitter and receiver are equipped with multiple antennas and the transmitter selects a finite number of transmit antennas. The TAS protocol, considered here, selects the antennas which observe the strongest channel gains to the receiver. Our results show that in the asymptotic regime, the gap between the expected mutual information and the upper bound derived by applying Jensen's inequality, remains constant in terms of N_t . Using the large-system approximation, we further investigate the asymptotic channel hardening property for this setup. For the special case of single-antenna receiver, our result reduces to the approximation reported in the literature.

Notation: Scalars, vectors and matrices are represented with non-bold, bold lower case and bold upper case letters, respectively. \mathbf{H}^H indicates the Hermitian of \mathbf{H} , and \mathbf{I}_N is the $N \times N$ identity matrix. The determinant of \mathbf{H} and euclidean norm of \mathbf{x} are denoted by $|\mathbf{H}|$ and $\|\mathbf{x}\|$. $\log(\cdot)$ and $\ln(\cdot)$ indicate the binary and natural logarithm, and $\mathbb{E}\{\cdot\}$ is the expectation operator. The beta distribution with the shape parameters α and β is denoted by $\text{Beta}(\alpha, \beta)$.

II. PROBLEM FORMULATION

We consider a Gaussian MIMO channel in which the transmitter and receiver are equipped with N_t and N_r antennas, re-

spectively. The transmitter selects L_t transmit antennas based on the information provided through a rate-limited return channel. For this setup, we investigate the input-output mutual information when the number of transmit antennas grows large.

A. System Model

The received signal by the receiver at each time interval is denoted by $\mathbf{y}_{N_r \times 1}$ and reads

$$\mathbf{y} = \sqrt{\rho} \mathbf{H} \mathbf{x} + \mathbf{n}, \quad (1)$$

where ρ denotes the average SNR at each receive antenna, $\mathbf{n}_{N_r \times 1}$ is circularly symmetric zero-mean complex Gaussian noise with unit variance, i.e., $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, $\mathbf{x}_{N_t \times 1}$ identifies the transmit signal with the power constraint $\mathbb{E} \mathbf{x}^H \mathbf{x} \leq 1$, and \mathbf{H} denotes an $N_r \times N_t$ independent and identically distributed (i.i.d.) unit-variance Rayleigh fading channel. It is assumed that the Channel State Information (CSI) is available only at the receiver side.

B. TAS Protocol

The transmitter, at each time interval, selects the L_t strongest channels by employing the TAS protocol \mathcal{S} . To illustrate the protocol, let \mathbf{h}_j denote the j th column vector of \mathbf{H} for $j \in \{1, \dots, N_t\}$. Moreover, represent the index set of order statistics from the arranging of vectors $\|\mathbf{h}_j\|^2$ in decreasing order of magnitude by $\{w_1, \dots, w_{N_t}\}$, i.e.,

$$\|\mathbf{h}_{w_1}\|^2 \geq \|\mathbf{h}_{w_2}\|^2 \geq \dots \geq \|\mathbf{h}_{w_{N_t}}\|^2. \quad (2)$$

At each time interval, the receiver informs the transmitter about the set $\{w_1, \dots, w_{L_t}\}$ through a rate-limited return channel. The transmitter, then, selects the corresponding antennas.

C. Input-Output Mutual Information

Suppose that independent Gaussian symbols are transmitted on the selected antennas. In this case, the mutual information between the input vector \mathbf{x} and the output \mathbf{y} denoted in (1), for a given realization of \mathbf{H} , is written as [8]

$$\mathcal{I}(\mathbf{H}; \rho \mathbf{Q}) := \log |\mathbf{I}_{N_r} + \rho \mathbf{H} \mathbf{Q} \mathbf{H}^H| \quad (3)$$

where \mathbf{Q} is an $N_t \times N_t$ diagonal matrix with nonzero diagonal entries at the indices $\{w_1, \dots, w_{L_t}\}$ and zero at the rest. When the power is uniformly allocated among the selected antennas, the nonzero entries of \mathbf{Q} equal to L_t^{-1} , and therefore, the input-output mutual information reduces to

$$\mathcal{I}_{\mathcal{S}} := \mathcal{I}(\tilde{\mathbf{H}}; \frac{\rho}{L_t} \mathbf{I}_{L_t}) = \log |\mathbf{I}_{N_r} + \frac{\rho}{L_t} \tilde{\mathbf{H}} \tilde{\mathbf{H}}^H| \quad (4)$$

where $\tilde{\mathbf{H}}$ is an $N_r \times L_t$ matrix describing the effective channel between the transmitter and the receiver, and constructed from \mathbf{H} by collecting the columns which correspond to the selected antennas, i.e., $\tilde{\mathbf{H}} = [\mathbf{h}_1, \dots, \mathbf{h}_{L_t}]$, where $\{\mathbf{h}_1, \dots, \mathbf{h}_{L_t}\}$ is a permutation of $\{\mathbf{h}_{w_1}, \dots, \mathbf{h}_{w_{L_t}}\}$. For a given realization of \mathbf{H} , $\mathcal{I}_{\mathcal{S}}$ upper bounds achievable transmit rates, since the CSI is only available at the receiver. Thus, one can define $\mathcal{I}_{\mathcal{S}}$ to be the maximum achievable rate under the TAS protocol \mathcal{S} .

III. ASYMPTOTIC HARDENING PROPERTY UNDER TAS

When the transmitter employs all the transmit antennas with uniform power allocation, the input-output mutual information is determined by letting $\mathbf{Q} = N_t^{-1} \mathbf{I}_{N_t}$ in (3). In this case, as N_t grows large with a fixed N_r , $N_t^{-1} \mathbf{H} \mathbf{H}^H$ converges to \mathbf{I}_{N_r} due to the law of large numbers, and thus, the mutual information converges to $N_r \log(1 + \rho)$ in large limits. This property is known as the ‘‘asymptotic hardening’’ property and has been rigorously justified for i.i.d. Rayleigh fading channels when the number of antennas at one side grows large [7].

A. Channel Hardening under TAS

Considering the TAS protocol \mathcal{S} , the channel matrix is a finite collection of order statistics obtained from arranging the magnitude of i.i.d. channel vectors. The case differs from which considered in [7], and therefore, the asymptotic analyses therein can not be extended. For this case, one can write

$$\mathcal{I}_{\mathcal{S}} = \sum_{\ell=1}^L \log \left(1 + \frac{\rho}{L_t} \lambda_{\ell} \right) \quad (5)$$

where $L := \min \{L_t, N_r\}$, and λ_{ℓ} identifies the ℓ th eigenvalue of $\mathbf{J}_{L \times L}$ defined as

$$\mathbf{J} = \begin{cases} \tilde{\mathbf{H}}^H \tilde{\mathbf{H}}, & \text{if } L = L_t \\ \tilde{\mathbf{H}} \tilde{\mathbf{H}}^H, & \text{if } L = N_r. \end{cases} \quad (6)$$

Consequently, Jensen’s inequality [9] suggests that

$$\mathcal{I}_{\mathcal{S}} \leq^* L \log \left(1 + \frac{\rho}{L_t L} \sum_{\ell=1}^L \lambda_{\ell} \right) \quad (7a)$$

$$= L \log \left(1 + \frac{\rho}{L_t L} \text{Tr} \{ \mathbf{J} \} \right) \quad (7b)$$

where the equality in $*$ holds when $L = 1$. Considering the upper bound in (7b), $\text{Tr} \{ \mathbf{J} \}$ is the sum of limited number of order statistics taken from a large arranging set. The sum is known as a trimmed sum in the literature and is shown to converge to a Gaussian random variable in distribution, when N_t grows large. This property of trimmed sums justifies the asymptotic hardening property for the upper bound. We approve that the property extends to $\mathcal{I}_{\mathcal{S}}$ as well by approximating the mutual information from the upper bound geometrically.

B. Main Result

Proposition 1 approximates $\mathcal{I}_{\mathcal{S}}$ for large N_t with a Gaussian random variables whose mean lies within a fixed gap below the mean of the upper bound in (7b) and whose variance converges to zero as $N_t \uparrow \infty$.

Proposition 1. Consider the TAS protocol \mathcal{S} . For large N_t , $\mathcal{I}_{\mathcal{S}}$ is approximated with a Gaussian random variable with mean η and variance σ^2 where

$$\eta = L \left[\log \left(1 + \frac{\rho \eta_t}{L_t L} \right) - \frac{(L-1) \rho^2 \eta_t^2 \log e}{2M (L_t L + \rho \eta_t)^2} \right] \quad (8a)$$

$$\sigma^2 = \left[\frac{\xi L \rho}{L_t L + \rho \eta_t} \log e \right]^2 \sigma_t^2 \quad (8b)$$

where η_t and σ_t^2 are determined in Appendix A and read

$$\eta_t = N_r L_t \left[1 + \mathcal{O}(\ln \left(\frac{N_t}{L_t} \right)) \right] \quad (9a)$$

$$\sigma_t^2 = N_r L_t \left[N_r + 1 - \mathcal{O}\left(\frac{L_t^{N_r}}{N_t^{N_r}}\right) \right]. \quad (9b)$$

and ξ is given by

$$\xi = 1 - \frac{L_t L (L-1) \rho \eta_t}{M (L_t L + \rho \eta_t)^2}, \quad (10)$$

with $L := \min \{L_t, N_r\}$ and $M := \max \{L_t, N_r\}$.

Proof. The proof is sketched through the large-system analysis in Section V. The details, however, are left for the extended version of the paper.

Proposition 1 illustrates the asymptotic hardening property of the MIMO channel, under the TAS protocol \mathcal{S} . In fact, as N_t grows large, η_t grows proportionally large, and $\sigma_t^2 \ll \eta_t$. Consequently, $\sigma^2 \downarrow 0$ and η reads

$$\eta \rightarrow L \log \left(1 + \frac{\rho \eta_t}{L_t L} \right) - \frac{L(L-1)}{2M} \log e. \quad (11)$$

Comparing the mean and variance of \mathcal{I}_S with the upper bound derived by Jensen's inequality, one observes that

$$\lim_{N_t \uparrow \infty} \log \left(1 + \frac{\rho}{L_t L} \text{Tr} \{ \mathbf{J} \} \right) - \frac{1}{L} \mathcal{I}_S = \frac{L-1}{2M} \log e. \quad (12)$$

(12) states that the gap between \mathcal{I}_S and the upper bound given by Jensen's inequality remains fixed asymptotically; the property which indicates that both \mathcal{I}_S and the upper bound exhibit a same limiting fluctuation. Noting that $\mathbb{E} \{ \mathcal{I}_S \}$ grows large proportional to N_t , the upper bound proposed by Jensen's inequality can be considered as a robust measure describing the asymptotics of the input-output mutual information within a constant scalar.

Special Case. Our main result recovers the special case of $N_r = 1$ studied in [6]. In fact, in this case $L = 1$, and thus, $\xi = 1$ which results in same η and σ^2 reported in [6]. The authors in [6], moreover, considered the absolute value of the asymptotic Gaussian random variable in Proposition 1 to be the approximation, in order to avoid approximating negative mutual information. The final expression, however, does not differ from the Gaussian random variable significantly, since probability of \mathcal{I}_S being approximated with a negative value is almost zero for large N_t .

IV. NUMERICAL RESULTS

For sake of comparison, the empirical cumulative distributions, and the corresponding approximations are demonstrated in Fig. 1 for various number of transmit antennas. The empirical distributions are obtained using 20000 channel realizations. Through our simulations, we assume $L_t = 16$, $N_r = 8$ and $\rho = 0$ dB. As Fig. 1 illustrates, the approximations, given by Proposition 1, meet the empirical distributions even within a finite number of antennas. In fact, although our analyses

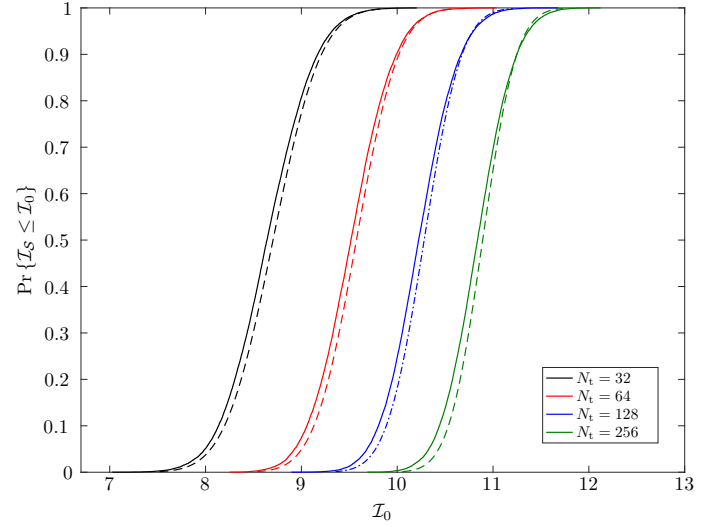


Fig. 1: Comparison of empirical cumulative distribution of \mathcal{I}_S and approximated distribution given by Proposition 1 for various number of transmit antennas. The solid and dashed lines indicate the approximated and empirical distribution, respectively. SNR is set to be $\rho = 0$ dB, $N_r = 8$ and $L_t = 16$.

considered the system in the large limit, the simulations show the validity of the results even in non-asymptotic scenarios. Proposition 1, enables us to accurately approximate diverse performance measures on fading channels, such as ergodic and outage capacity which we briefly address in the sequel.

A. Ergodic Capacity

For the setup illustrated in Section II-A, the ergodic capacity is defined as the maximum average achievable transmission rate and given by taking the expectation of the input-output mutual information \mathcal{I}_S . Using Proposition 1, the ergodic capacity of the channel \mathbf{H} , under the TAS protocol \mathcal{S} , is approximated by η . Fig. 2 shows the ergodic capacity as a function of average receive SNR per antenna for various number of selected antennas, when the transmitter and receiver are equipped with $N_t = 128$ and $N_r = 16$ antennas, respectively. The numerical results show that, for the given range of SNRs, the approximation tracks the simulation results with maximum of approximately 2% deviation.

B. Outage Capacity

In slow fading scenarios, where the channel does not fluctuate significantly within the transmission interval, the ergodic capacity cannot describe the real transmission limit on the channel. In this case, one may consider the outage capacity $\mathcal{R}_{\text{out}}(p_{\text{out}})$ which for a given outage probability p_{out} reads

$$\Pr \{ \mathcal{I}_S \leq \mathcal{R}_{\text{out}}(p_{\text{out}}) \} = 1 - p_{\text{out}}. \quad (13)$$

Fig. 3 plots the 10% outage capacity, i.e., $p_{\text{out}} = 0.1$, versus the number of selected antennas for different N_r , assuming $N_t = 128$ and $\rho = 0$ dB. As the figure illustrates, for the given numbers of selected antennas, Proposition 1 meets the numerical simulations with approximately 1.5% deviation at most.

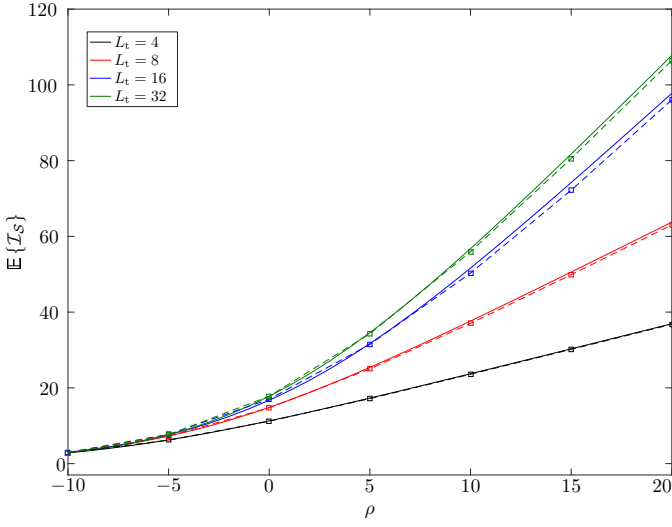


Fig. 2: Ergodic capacity as a function of SNR. The solid lines indicate the approximated ergodic capacity given by Proposition 1, and the dashed lines are plotted via numerical simulations. The number of transmit and receive antennas are set to be $N_t = 128$ and $N_r = 16$.

V. LARGE-SYSTEM ANALYSIS

In this section, we briefly sketch the proof of Proposition 1. Due to the lack of space, we omit the detailed derivations here and give them in the extended version of the manuscript. Our derivations mainly follow two steps:

- (A) An approximation for the input-output mutual information of a Gaussian MIMO channel is calculated. Using the approximation, \mathcal{I}_S is given in terms of a trimmed sum.
- (B) The asymptotic properties of trimmed sums, as well as random matrices, are employed to determine the statistics of the selected channel in the large-system limit.

A. Approximating the Mutual Information

Considering the effective channel $\tilde{\mathbf{H}}$, we derive an approximation for the input-output mutual information in terms of the first and second order statistics of \mathbf{J} defined in (6). Starting from (5), let us define the scalars i_ℓ for $\ell \in \{1, \dots, L\}$ to be

$$i_\ell = \log \left(1 + \frac{\rho}{L_t} \lambda_\ell \right). \quad (14)$$

The tuple (μ, c) then denotes the centroid of the set

$$\mathcal{A} = \{(\lambda_\ell, i_\ell) \mid \forall \ell \in \{1, \dots, L\}\}, \quad (15)$$

and reads

$$c = \frac{1}{L} \sum_{\ell=1}^L i_\ell = \frac{1}{L} \mathcal{I}_S, \quad (16a)$$

$$\mu = \frac{1}{L} \sum_{\ell=1}^L \lambda_\ell = \frac{1}{L} \text{Tr} \{\mathbf{J}\}. \quad (16b)$$

Therefore, the upper bound in (7b) reduces to

$$c \leq \log \left(1 + \frac{\rho}{L_t} \mu \right). \quad (17)$$

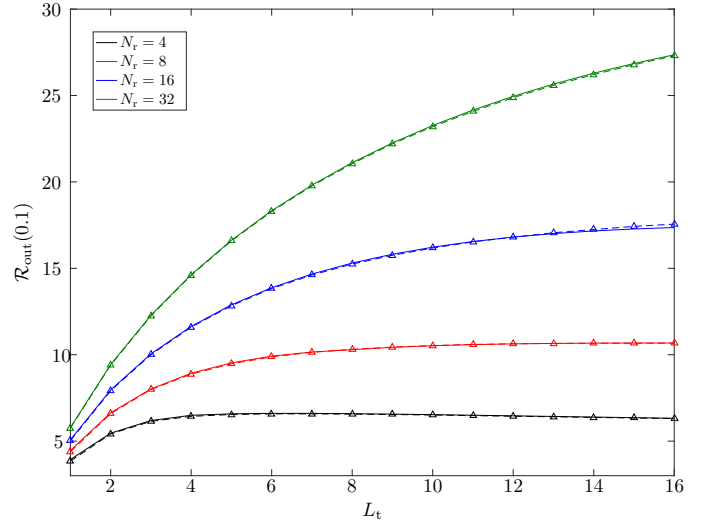


Fig. 3: 10% outage capacity in terms of number of selected antennas. The solid and dashed lines respectively denote the approximations, and the numerical simulations for $N_t = 128$ at $\rho = 0$ dB.

(17) suggests a geometric approximation for c which we intuitively illustrate here.

Consider the mutual information curve $\log(1 + \rho L_t^{-1} x)$. By deviating from μ on the curve with a certain step size δ , the tuples $(\mu - \delta, i_-)$ and $(\mu + \delta, i_+)$ are obtained where

$$i_- = \log \left(1 + \frac{\rho}{L_t} (\mu - \delta) \right), \quad (18a)$$

$$i_+ = \log \left(1 + \frac{\rho}{L_t} (\mu + \delta) \right). \quad (18b)$$

Due to continuity and concavity of the mutual information curve, the centroid of the line connecting $(\mu - \delta, i_-)$ and $(\mu + \delta, i_+)$ lies under (μ, c) for some choices of δ . The point, moreover, can be arbitrarily close to (μ, c) , if δ is set properly; see Fig. 4. As the result, \mathcal{I}_S can be approximated as

$$\mathcal{I}_S \approx \frac{L}{2} (i_- + i_+). \quad (19)$$

Using the polynomial expansion, (19) is written as

$$\mathcal{I}_S = L \left[\log \left(1 + \frac{\rho}{L_t} \mu \right) - \frac{\kappa^2 \delta^2}{2} \log e \right] + \mathcal{O}(\kappa^4 \delta^4) \quad (20)$$

where κ is defined to be

$$\kappa := \frac{\rho}{L_t + \rho \mu}. \quad (21)$$

At this point, we need to determine δ . To do so, we note that the input-output mutual information in (4), can be expanded in an alternative way. Define Δ to be an $L \times L$ matrix, such that $\mathbf{J} = \mu \mathbf{I}_L + \Delta$. Consequently, one can write

$$\mathcal{I}_S = L \log \left(1 + \frac{\rho}{L_t} \mu \right) + \log |\mathbf{I}_L + \kappa \Delta| \quad (22)$$

where κ is defined in (21). We expand the second term in the right hand side (r.h.s.) of (22) by evaluating a polynomial expansion for the determinant of a perturbed identity matrix.

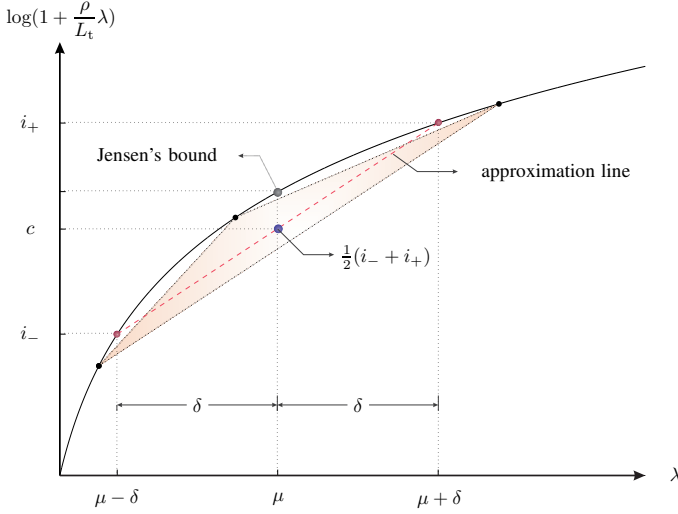


Fig. 4: Geometric approximation of $\mathcal{I}(\mathcal{S})$. By choosing δ properly, the approximation line at $\lambda = \mu$ takes a value close to the centroid of the polytope.

Lemma 1. For κ in a small vicinity of zero, $|\mathbf{I} + \kappa \Delta|$ reads

$$|\mathbf{I} + \kappa \Delta| = 1 + \kappa \text{Tr}\{\Delta\} + \frac{\kappa^2}{2} [\text{Tr}\{\Delta\}^2 - \text{Tr}\{\Delta^2\}] + \mathcal{O}(\kappa^3). \quad (23)$$

Proof. Starting from the identity $|e^{\kappa \Delta}| = e^{\kappa \text{Tr}\{\Delta\}}$, the proof is concluded after some lines of derivations.

Using Lemma 1, (22) reduces to

$$\begin{aligned} \mathcal{I}_S &= L \log \left(1 + \frac{\rho \mu}{L_t} \right) + \log \left(1 - \frac{\kappa^2}{2} \text{Tr}\{\Delta^2\} + \mathcal{O}(\kappa^3) \right) \\ &\stackrel{*}{=} L \log \left(1 + \frac{\rho \mu}{L_t} \right) - \frac{\kappa^2}{2} \text{Tr}\{\Delta^2\} \log e + \mathcal{O}(\kappa^3) \end{aligned} \quad (24)$$

where \star follows from taking the assumption that $\kappa \downarrow 0$ as N_t grows large. We later show that the assumption holds, since μ grows proportional to N_t . By letting the r.h.s. of (20) and (24) to be equal, δ is determined as

$$\delta^2 = \frac{\text{Tr}\{\Delta^2\}}{L} = \frac{\text{Tr}\{\mathbf{J}^2\}}{L} - \frac{\text{Tr}\{\mathbf{J}\}^2}{L^2}. \quad (25)$$

Substituting (25) in (19), we conclude the following lemma.

Lemma 2. The input-output mutual information for any realization of \mathbf{H} is approximated as

$$\mathcal{I}_S \approx \frac{L}{2} \left[\log \left(1 + \frac{\rho(\mu - \delta)}{L_t} \right) + \log \left(1 + \frac{\rho(\mu + \delta)}{L_t} \right) \right] \quad (26)$$

where $\mu := \text{Tr}\{\mathbf{J}\}/L$ and δ is given as in (25).

B. Asymptotics of \mathcal{I}_S

The approximation proposed in Lemma 2 enables us to investigate the input-output mutual information in the large-system limit. Consider the TAS protocol \mathcal{S} . As it is indicated in (2), $\|\mathbf{h}_{w_\ell}\|^2$, for $\ell \in \{1, \dots, N_t\}$, are order statistics in decreasing order; therefore, $\text{Tr}\{\mathbf{J}\} = \sum_{\ell=1}^{L_t} \|\mathbf{h}_{w_\ell}\|^2$ is the sum of L_t first order statistics. In the context of order statistics, this sum is known as a trimmed sum, and shown to converge

to a Gaussian random variable in distribution, when the size of arranging set, i.e., N_t , tends to infinity [10]. In [11], the author determined the mean and variance of the asymptotic trimmed sum in terms of the distribution of the arranged random variables. Although the analytic result in [11] is determined in the asymptotic regime, the numerical results show that the distribution of a trimmed sum converges relatively fast to the limiting distribution. Therefore, the asymptotic result provides an efficient and robust approximation of the exact behavior of a large scale, but not infinite length, trimmed sum [10].

Considering the above discussion, for large N_t , $\text{Tr}\{\mathbf{J}\}$ can be approximated with the asymptotic distribution given in [11], when the distribution of the arranged variables, i.e., $\|\mathbf{h}_{w_\ell}\|^2$, is set to be chi-square with $2N_r$ degrees of freedom. In this case, by keeping L_t fixed, the accuracy of the approximation increases, when N_t grows larger. Using the main theorem of [11], $\text{Tr}\{\mathbf{J}\}$ is approximated with t , where $t \sim \mathcal{N}(\eta_t, \sigma_t^2)$. The exact values of η_t and σ_t^2 are determined in Appendix A. Using the large-system approximation, it is shown that the mean and variance for large N_t are as in (9a) and (9b).

Remark. As $\text{Tr}\{\mathbf{J}\} > 0$, the approximation of $\text{Tr}\{\mathbf{J}\}$ with a Gaussian random variable clearly fails for some realizations of \mathbf{J} , since t can take negative values. This is a direct result of the fact that t is only a large-system approximation of $\text{Tr}\{\mathbf{J}\}$. As N_t grows large, $\text{Pr}\{t < 0\}$ converges to zero, and therefore, the approximation becomes more accurate.

As we consider N_t to be significantly large, one can see that $\kappa \ll 1$ under the TAS protocol \mathcal{S} . Substituting the large-system approximation of $\text{Tr}\{\mathbf{J}\}$ in Lemma 2, μ is considered to be normally distributed around $L^{-1}\eta_t$ with variance $L^{-2}\sigma_t^2$. Therefore, using the polynomial expansion, the approximation in Proposition 2 reduces to

$$\mathcal{I}_S \approx L \log \left(1 + \frac{\rho}{L_t L} t \right) - \frac{L}{2} \kappa^2 \delta^2 \log e. \quad (27)$$

The first term in the r.h.s. of (27) is further expanded as

$$\log \left(1 + \frac{\rho}{L_t L} t \right) \stackrel{\dagger}{=} \log \left(1 + \frac{\rho \eta_t}{L_t L} \right) + \frac{\rho(t - \eta_t)}{L_t L + \rho \eta_t} \log e + \epsilon_{N_t} \quad (28)$$

where ϵ_{N_t} converges to zero as N_t tends to infinity. \dagger comes from the polynomial expansion of $\log(1+x)$ at $x=0$, in which the higher order terms are dropped. Using the same argument as for (27)-(28), it is shown that, in the large limit, κ^2 reads

$$\kappa^2 = \frac{L^2 \rho^2}{(L_t L + \rho \eta_t)^2} \left[1 - 2 \frac{\rho(t - \eta_t)}{L_t L + \rho \eta_t} \right] + \epsilon_{N_t}. \quad (29)$$

Although δ in general finds a complicated distribution, in the large limit, it could be approximated in a straightforward form in terms of t . To show that, consider \mathbf{J} defined in (6). Thus,

$$\text{Tr}\{\mathbf{J}^2\} = \sum_{\ell=1}^{L_t} \left[\|\mathbf{h}_{w_\ell}\|^4 + \|\mathbf{h}_{w_\ell}\|^2 \sum_{\substack{k=1 \\ k \neq \ell}}^{L_t} \|\mathbf{h}_{w_k}\|^2 \cos^2 \theta_{\ell,k} \right] \quad (30)$$

where $\theta_{\ell,k}$ denotes the Hermitian angle between \mathbf{h}_ℓ and \mathbf{h}_k and is defined as

$$\theta_{\ell,k} = \cos^{-1} \frac{|\mathbf{h}_{w_\ell}^H \mathbf{h}_{w_k}|}{\|\mathbf{h}_{w_\ell}\| \|\mathbf{h}_{w_k}\|}. \quad (31)$$

By the same approach as the one taken for approximating the distribution of $\text{Tr}\{\mathbf{J}\}$, the trace of any principle submatrix of \mathbf{J} can be determined in the large limit. In fact, by using the main theorem of [11], it is shown that the sum of any subset \mathcal{S} of order statistics $\|\mathbf{h}_{w_\ell}\|^2$ is approximately normally distributed around $|\mathcal{S}|L_t^{-1}\eta_t$ with a variance whose tends to zero as N_t grows large, where $|\mathcal{S}| \leq L_t$ indicates the size of \mathcal{S} . Consequently, $\|\mathbf{h}_{w_\ell}\|^2 = L_t^{-1}\text{Tr}\{\mathbf{J}\} + \alpha_\ell$ for any $\ell \in \{1, \dots, L_t\}$ where α_ℓ is a zero-mean random variable with variance converging to zero. Substituting in (30),

$$\text{Tr}\{\mathbf{J}^2\} = \frac{\text{Tr}\{\mathbf{J}\}^2}{L_t} + \frac{\text{Tr}\{\mathbf{J}\}^2}{L_t^2} \sum_{\substack{\ell,k=1 \\ k \neq \ell}}^{L_t} \cos^2 \theta_{\ell,k} + \epsilon_{N_t}, \quad (32)$$

where ϵ_{N_t} tends to zero as N_t grows large. Considering the TAS protocol \mathcal{S} , the ordering in (2) considers only the magnitude of \mathbf{h}_j . Therefore, the distribution of $\theta_{\ell,k}$ is same as the distribution of the Hermitian angles between the column vectors of an i.i.d. complex Gaussian channel. For this case, it has been reported that $\theta_{\ell,k}$, for a given ℓ , are independent [12]. Moreover, the distribution of $\cos^2 \theta_{\ell,k}$ has been shown to be Beta(1, $N_r - 1$), see Appendix C of [12]. Thus,

$$\text{Tr}\{\mathbf{J}^2\} = \frac{\text{Tr}\{\mathbf{J}\}^2}{L_t} [1 + (L_t - 1)\beta] + \epsilon_{N_t} \quad (33)$$

where β is defined as $\beta := L_t^{-1} \sum_{\ell=1}^{L_t} \tilde{\beta}_\ell$ with

$$\tilde{\beta}_\ell := \frac{1}{L_t - 1} \sum_{k=1, k \neq \ell}^{L_t} \cos^2 \theta_{\ell,k}. \quad (34)$$

The summand in r.h.s. of (34) is a sequence of independent beta distributed random variables. Therefore, the distribution of $\tilde{\beta}_\ell$ is given by $L_t - 1$ times convolution of Beta(1, $N_r - 1$), and then normalizing correspondingly. Using the properties of beta distribution, it is then shown that even for finite L_t and N_r , $\tilde{\beta}_\ell$ are approximately distributed normally around N_r^{-1} with a variance significantly smaller than N_r^{-1} . Consequently, β is approximately a Gaussian random variable with mean N_r^{-1} . As $\tilde{\beta}_\ell$ are in general dependent, the variance of β is not simply written as the sum of the variances; however, one can upper bound the variance by considering the extreme case of full dependency. Therefore, $\text{Tr}\{\mathbf{J}^2\}$ is written as

$$\text{Tr}\{\mathbf{J}^2\} = \left[\frac{N_r + L_t - 1}{N_r L_t} + \chi \right] \text{Tr}\{\mathbf{J}\}^2 + \epsilon_{N_t}, \quad (35)$$

where χ is a zero-mean random variable with approximately Gaussian distribution whose variance is relatively small compared to $N_r^{-1} L_t^{-1} (N_r + L_t - 1)$. Defining $M := \max\{L_t, N_r\}$, the scalar δ in Lemma 2 reads

$$\delta^2 = \left[\frac{L - 1}{ML^2} + \frac{\chi}{L} \right] \text{Tr}\{\mathbf{J}\}^2 + \epsilon_{N_t}. \quad (36)$$

Using the large-system approximation for $\text{Tr}\{\mathbf{J}\}$, i.e., t , and taking the same steps as (28), we have

$$\delta^2 = \left[\frac{L - 1}{ML^2} + \frac{\chi}{L} \right] \left[1 + 2 \frac{t - \eta_t}{\eta_t} \right] \eta_t^2. \quad (37)$$

Finally by substituting (28), (29) and (37) in (27), and after some lines of derivations Proposition 1 is concluded.

APPENDIX A ASYMPTOTICS OF $\text{Tr}\{\mathbf{J}\}$

In order to find the exact asymptotic characteristics of the random variable $\text{Tr}\{\mathbf{J}\}$ in Section V-B, we invoke the result reported in [11]. Using the main theorem of [11], η_t reads

$$\eta_t = N_r [L_t + N_t f_{N_r+1}(u)] \quad (38)$$

where $f_{N_r}(\cdot)$ denotes the chi-square probability density function with $2N_r$ degrees of freedom and mean N_r ,

$$f_{N_r}(x) = \frac{1}{(N_r - 1)!} \begin{cases} x^{N_r-1} e^{-x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (39)$$

and u is the solution of the equation

$$\int_u^\infty f_{N_r}(x) dx = \frac{L_t}{N_t}. \quad (40)$$

Moreover, σ_t^2 is determined as

$$\sigma_t^2 = (L_t u - \eta_t)^2 \left(\frac{1}{L_t} - \frac{1}{N_t} \right) - \frac{\eta_t^2}{L_t} + \Xi_t \quad (41)$$

where the non-negative scalar Ξ_t is defined as

$$\Xi_t = N_r (N_r + 1) [L_t + N_t f_{N_r+1}(u) + N_t f_{N_r+2}(u)]. \quad (42)$$

REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.
- [3] A. F. Molisch, M. Z. Win, Y.-S. Choi, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Trans. on Wireless Communications*, vol. 4, no. 4, pp. 1759–1772, 2005.
- [4] D. Bai, P. Mitran, S. S. Ghassemzadeh, R. R. Miller, and V. Tarokh, "Channel hardening and the scheduling gain of antenna selection diversity schemes," in *IEEE Int. Symp. on Inf. Theory*, pp. 1066–1070, 2007.
- [5] P. Hesami and J. N. Laneman, "Limiting behavior of receive antennae selection," in *45th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2011.
- [6] H. Li, L. Song, and M. Debbah, "Energy efficiency of large-scale multiple antenna systems with transmit antenna selection," *IEEE Trans. on Communications*, vol. 62, no. 2, pp. 638–647, 2014.
- [7] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.
- [8] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecomm.*, vol. 10, no. 6, pp. 585–595, 1999.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2012.
- [10] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*, Siam, vol. 54, 1992.
- [11] S. M. Stigler, "The asymptotic distribution of the trimmed mean," *The Annals of Statistics*, pp. 472–477, 1973.
- [12] R. R. Müller, "Multiuser receivers for randomly spread signals: Fundamental limits with and without decision-feedback," *IEEE Trans. on Inf. Theory*, vol. 47, no. 1, pp. 268–283, 2001.